THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1993 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 268, No. 19, Issue of July 5, pp. 14090-14095, 1993
Printed in U.S.A.

Similar Alternative Splicing Events Generate Two Sarcoplasmic or Endoplasmic Reticulum Ca-ATPase Isoforms in the Crustacean *Artemia franciscana* and in Vertebrates*

(Received for publication, February 8, 1993)

Ricardo Escalante‡ and Leandro Sastre§

From the Instituto de Investigaciones Biomédicas del CSIC, Calle/Arturo Duperier, 4, 28029 Madrid, Spain

A second isoform of the *Artemia franciscana* sarcoplasmic or endoplasmic reticulum Ca-ATPase has been identified through the isolation of cDNA clones. This isoform differs from the previously identified one only at the C-terminal end of the protein. The last 6 amino acids of the former isoform change to 30 hydrophobic amino acids in the newly identified isoform that have the potentiality of being an additional transmembrane domain. The two *A. franciscana* isoforms are highly homologous to the two isoforms coded by the mammalian and bird SERCA2 gene, except that the C-terminal extensions share their hydrophobic character but have no significant amino acid homology. The isolation of genomic clones coding for this region of the gene shows that both isoforms arise from the same gene by alternative splicing. The donor splicing site of the penultimate exon can either be recognized and fused to the last exon, giving rise to the mRNA coding for the shortest protein, or remain unrecognized, in which case a polyadenylation site is recognized before the last exon of the gene and the mRNA coding for the largest protein is originated. The alternative splicing used to originate the two isoforms is similar in *Artemia* and vertebrates. The conservation of the alternative splicing between species so distant in evolution suggests an important physiological role for the existence of the two isoforms of the protein.

Sarcoplasmic or endoplasmic reticulum Ca-ATPases (SERCA)[1] are the enzymes responsible for the active transport of calcium from the cytoplasm to the sarcoplasmic reticulum of muscular cells or the endoplasmic reticulum of non-muscular cells after cell activation. Three SERCA genes have been described in mammals and birds and have been named SERCA1, -2, and -3 (Brandl *et al.*, 1986; Lytton and MacLennan, 1988; Burk *et al.*, 1989). These genes code for highly homologous proteins that differ in their tissue specificity of expression. SERCA1 is expressed in fast twitch skeletal muscles (Brandl *et al.*, 1987), SERCA2 is expressed in slow twitch, smooth, and cardiac muscles and in non-muscular cells (Lytton *et al.*, 1989; Gunteski-Hamblin *et al.*, 1988), and SERCA3 is expressed in some muscular and non-muscular cells (Burk *et al.*, 1989).

Besides the existence of three genes, additional diversity is obtained by alternative splicing mechanisms in birds and mammals. The SERCA1 primary transcript can be processed in two alternative ways; the penultimate exon is retained in adult muscles but is spliced out in neonatal animals. Because of this alternative splicing, the adult isoform codes for a protein that presents a glycine as C-terminal amino acid instead of the last 8 amino acids of the neonatal isoform (Brandl *et al.*, 1987).

The SERCA2 primary transcript is also spliced differently depending on the tissue of expression (Lytton *et al.*, 1989; Gunteski-Hamblin *et al.*, 1988; Lytton and MacLennan, 1988). The donor splicing site of the penultimate exon is recognized in cardiac muscle cells, where it is fused to the last exon of the gene. In smooth muscle and non-muscle cells, this donor site is not recognized and transcription is stopped at a polyadenylation site located before the last exon of the gene. This alternative splicing originates two mRNAs that differ completely at their 3′ end, including the last amino acids of the protein. The last 4 C-terminal amino acids of the cardiac isoform are changed to 49 hydrophobic amino acids in the smooth muscle and non-muscular isoform (Lytton and MacLennan, 1988). The existence of similar alternative splicing events of the SERCA2 gene, giving rise to highly homologous isoforms, has been also described in birds (Campbell *et al.*, 1991). The function of this hydrophobic region is unknown at the present time.

The SERCA genes have been much less studied in invertebrates, although cDNA clones have been isolated coding for highly homologous proteins from *Drosophila* (Magyar and Váradi, 1990) and the crustacean *Artemia* (Palmero and Sastre, 1989). Chromosomal mapping and Southern and Northern blot hybridization experiments suggest the existence of a single gene coding for sarcoplasmic or endoplasmic reticulum Ca-ATPases in *Drosophila*. In contrast, Northern blot hybridization experiments showed the existence of two mRNAs coding for *Artemia* sarcoplasmic or endoplasmic reticulum Ca-ATPases that are differently regulated during embryonic development. We wanted to study if these two mRNAs are encoded by different genes or if they are the consequence of differential splicing of a unique primary transcript, as described in vertebrates. By isolating cDNA clones coding for both mRNAs, we show in this article that they are originated by differential splicing of a unique transcript. Par-

* This work was supported by Grant PB87-0208 from the Dirección General de Investigación Científica y Técnica. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) X72713.

‡ Recipient of a predoctoral fellowship from the Comunidad Autónoma de Madrid.

§ To whom correspondence should be addressed. Tel.: 34-1-5854626; Fax: 34-1-5854587.

[1] The abbreviations used are: SERCA, sarcoplasmic or endoplasmic reticulum Ca-ATPase; SSC, 0.15 M NaCl, 0.015 M trisodium citrate; kb, kilobase(s).

tial genomic clones coding for this gene have also been isolated, showing that the alternative splicing that originates the two mRNAs in *Artemia* is identical to that used in the vertebrate SERCA2 gene.

MATERIALS AND METHODS

Isolation of cDNA Clones—A cDNA library was synthesized from *Artemia franciscana* encysted gastrula mRNA (San Francisco Bay Brand, Newark, CA) in the λ vector ZAP-II, utilizing the ZAP cDNA synthesis kit from Stratagene. A total of 125,000 independent clones were obtained, and the library was plate-amplified once. A total of 150,000 independent clones from the amplified library were screened using the *Artemia* sarcoplasmic or endoplasmic reticulum Ca-ATPase clones pArtATCa151 and pArtATCa412 (Palmero and Sastre, 1989) as probe. Hybridization of the filters was made in 6 × SSC, 1% SDS, 50% formamide, 5 × Denhardt's, 100 μg/ml calf thymus DNA, and 10^6 cpm/ml probe at 42 °C for 15 h. Filters were washed three times for 30 min in 2 × SSC, 1% SDS at 65 °C. Clones pArATCa501, -505, -508, and -517 were isolated in this screening.

A second screening of a cDNA library made from 20-h-old nauplii mRNA in the plasmid vector pUC18 (Palmero and Sastre, 1989) was made under the same conditions using the 3′ end *Xba*I-*Xho*I fragment of the clone pArATCa501 to isolate the clone pArATCa604.

Isolation of Genomic Clones—A genomic library was constructed by partial digestion of *A. franciscana* DNA with MboI and ligation into the *Bam*HI site of the λ vector EMBL3. *Artemia* DNA was isolated from 20-h-old nauplii as described by Cruces *et al.* (1981). The ligated DNA was packaged *in vitro* using Gigapack II Gold packaging extracts from Stratagene, and 10^7 independent clones were obtained. A total of 750,000 clones from the unamplified library were screened using the cDNA clones pArATCa151 and -412 as probes. Hybridization and washing conditions were the same described for the isolation of cDNA clones. Clones coding for the 3′ end of the gene were selected between the positive phages by hybridization with the *Xba*I-*Xho*I fragment of the clones pArATCa501 and -505. Three clones hybridized to both of these two probes and were selected for further study (gArATCa2, -5, and -11).

RNA Hybridization—*Artemia* cryptobiotic embryos were developed for 6, 8, or 10 h according to Osuna and Sebastian (1980). Total RNA was prepared by the method of Sherman *et al.* (1986) except that homogenization was performing using a Polytron homogenizer and that the final precipitation of the RNA was made in 3 M LiCl at 4 °C, overnight. Fifteen micrograms of each RNA were electrophoresed in 1.5% agarose, 2.2 M formaldehyde gels, transferred to nylon membranes (Zeta Probe, Bio-Rad), and hybridized with the indicated probes. Hybridization and washing conditions were the same as those described for the isolation of cDNA clones.

DNA Sequencing—The nucleotide sequence of the DNAs was determined by the dideoxy chain-termination method as modified by Chen and Seeburg (1985) for sequencing double-stranded DNA. The different fragments of the cDNA or genomic clones required for sequencing were inserted in the plasmid vector pUC18. Either universal M13 primers or specific primers designed from previously established sequences were utilized. The *Taq* Dye Deoxy Terminator Cycle Sequencing kit and the model 373A Sequencer from Applied Biosystems were used for most of the sequencing experiments except a few where the T7-Sequencing-TM Kit from Pharmacia was utilized.

RESULTS

Isolation of cDNA Clones Coding for the Artemia SERCA 5.2-kb mRNA—The *Artemia* SERCA cDNA clones previously isolated hybridize with two mRNAs of 4.5 and 5.2 kb. These clones were all identical and were isolated from cDNA libraries made from embryos developed for 20 h or adult animals where the levels of expression of the 4.5-kb mRNA is much higher than that of the 5.2-kb mRNA (Palmero and Sastre, 1989). In cryptobiotic embryos (cysts) the levels of expression of both mRNAs is similar; therefore, we decided to construct a cDNA library from mRNA obtained from cysts to attempt the isolation of cDNA clones coding for the 5.2-kb mRNA. The screening of this library with two of the cDNA clones previously isolated (pArATCa151 and pArATCa412) under low stringency hybridization conditions yielded a number of cDNA clones, the restriction maps of which are represented in Fig. 1*A*. There were two types of cDNA clones; some clones presented a restriction map identical to that of the previously isolated cDNA clones (clones pArATCa505 and -508 in Fig. 1*A*), and other clones were identical to the previous ones in their 5′ region but were longer and presented a different restriction map in their 3′ ends (clones pArATCa501, -517, and -604 in Fig. 1*A*).

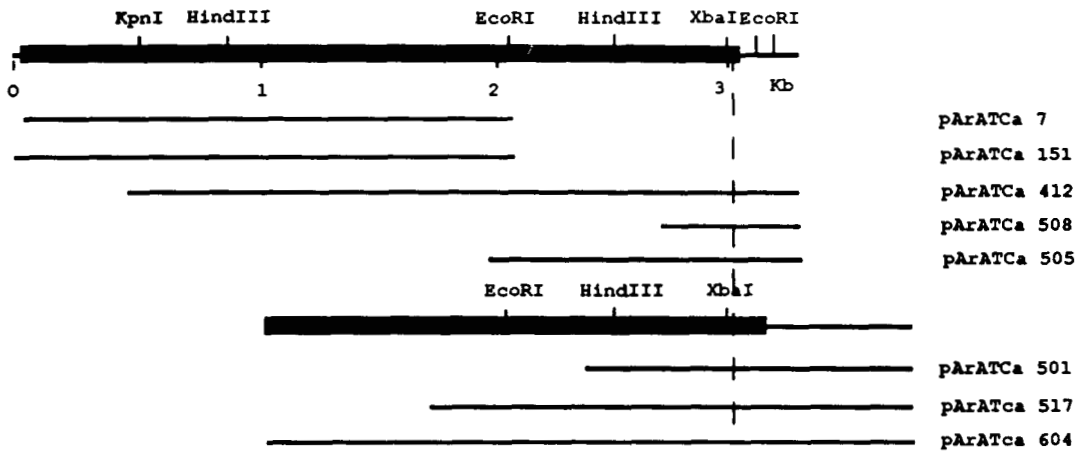
The nucleotide sequence of the longest cDNA clone of the second type (clone pArATCa604) was determined and is presented in Fig. 1*B*. The nucleotide sequence of this clone is identical to that of the previously characterized clones up to the nucleotide 1934 (corresponding to nucleotides 1167–3070 of the previously sequenced cDNAs), except for two conservative nucleotide changes and one non-conservative change (the A in nucleotide 382 changes to G, and the encoded Thr changes to Ala). The sequence differs completely from that of the other clones from nucleotide 1935 to the end of the clone. The divergent sequence includes the region coding for the last 6 amino acids of the previously characterized cDNAs. The clone pArATCa604 codes for 30 additional amino acids in the C-terminal region of the protein that are highly hydrophobic. The relation between the proteins encoded by these two *Artemia* cDNA clones is similar to the one observed for the two mammalian SERCA2 isoforms where the last 4 amino acids of one isoform are changed by 49 amino acids in the other isoform. The extension of 30 amino acids of the *Artemia* isoform do not show a significant homology to the 49 amino acids of the mammalian SERCA2b isoform; however, they share a marked hydrophobic character, as shown in Fig. 2, where the hydropathy profiles of the two *Artemia* and mammalian isoforms are shown.

To ascertain if the two types of cDNA clones corresponded to the 4.5- and 5.2-kb messengers, RNAs from *Artemia* embryos developed for 6, 8, or 10 h (*lanes 1*, *2*, and *3* in Fig. 3) were hybridized with DNA probes from either a region common to both types of cDNA clones (*probe A* in Fig. 3) or regions specific for each type of cDNA (*probes B* and *C* in Fig. 3). Probe A, from the cDNA clone pArATCa151, common to both types of cDNA clones, hybridized to the 4.5- and 5.2-kb mRNAs (Fig. 3*A*). In contrast, probe B, the *Xba*I-*Xho*I fragment of the clone pArATCa505, hybridized only with the 4.5-kb mRNA (Fig. 3*B*), and probe C, the *Xba*I-*Xho*I fragment of the clone pArATCa501, hybridized only with the 5.2-kb mRNA (Fig. 3*C*). These results show that the two types of cDNA clones isolated corresponded to each one of the two mRNAs observed previously.

Isolation of Genomic Clones Coding for the 3′ End of the Artemia SERCA Gene—The nucleotide sequence of the two types of cDNA clones suggested that they could have been originated by alternative splicing of a single primary transcript, as described for mammalian SERCA1 and SERCA2 genes (Brandl *et al.*, 1987; Lytton and MacLennan, 1988). To ascertain this possibility we have isolated genomic clones coding for the 3′ end of this gene. A genomic library was constructed in the λ EMBL3 vector using DNA from *Artemia* nauplii of the same batch used to construct the cDNA libraries. Screening of the genomic library with the cDNA clones pArATCa151 and -412 yielded a number of positive clones, three of which hybridized with DNA probes from the 3′ end of the clones pArATCa501 and -505 (*probes B* and *C* in Fig. 3) and were chosen for analysis.

The restriction map of these three clones is shown in Fig. 4. cDNA coding regions were localized and sequenced using specific oligonucleotides deduced from the cDNA clones. These experiments allowed the identification of three exons

A



B

```
   Asp Lys Thr Gly Thr Leu Thr Thr Asn Gln Met Ser Val Ser Arg Met Phe Val Phe Lys Asp Ile Pro Asp Asp Ala Ala    27
 1 GAC AAA ACT GGC ACA CTG ACA ACC AAC CAG ATG TCA GTT TCG AGG ATG TTC GTA TTC AAA GAC ATC CCA GAT GAT GCT GCT
   Pro Glu Leu Tyr Gln Phe Glu Leu Thr Gly Ser Thr Tyr Glu Pro Ile Gly Glu Thr Phe Met Gln Gly Gln Lys Ile Asn    54
 83 CCT GAA TTG TAT CAA TTC GAA CTA ACT GGA TCT ACC TAT GAA CCT ATT GGA GAG ACC TTT ATG CAA GGA CAA AAA ATC AAT
   Ala Ala Asp Tyr Asp Ala Val Lys Glu Ile Thr Thr Ile Cys Met Met Cys Asn Asp Ser Ala Ile Asp Phe Asn Glu Tyr    81
164 GCC GCT GAC TAT GAT GCT GTG AAA GAA ATC ACC ACA ATC TGT ATG ATG TGT AAC GAC TCA GCT ATT GAT TTT AAT GAG TAT
   Lys Gln Ala Phe Glu Lys Val Gly Glu Ala Thr Glu Thr Ala Leu Ile Val Leu Gly Glu Lys Leu Asn Pro Tyr Asn Leu   108
245 AAA CAA GCG TTT GAG AAG GTT GGA GAA GCC ACT GAA ACT GCG TTG ATT GTT CTT GGT GAA AAA CTG AAC CCA TAC AAT TTA
   Ser Lys Ala Gly Lys Asp Arg Arg Ser Ala Ala Leu Val Val Arg Glu Asp Met Asp Ala Arg Trp Lys Lys Glu Phe Thr   135
326 TCA AAG GCC GGA AAA GAC CGT CGT AGT GCC GCC CTT GTT GTA CGA GAA GAT ATG GAT GCT AGG TGG AAA AAG GAA TTT ACT
   Leu Glu Phe Ser Arg Asp Arg Lys Ser Met Ser Ser Tyr Cys Val Pro Leu Lys Ala Gly Leu Leu Ser Asn Gly Pro Lys   162
407 TTG GAG TTT TCC CGT GAC CGT AAA TCT ATG TCA TCG TAC TGT GTA CCA TTG AAA GCT GGA CTC CTG AGC AAT GGC CCC AAG
   Met Phe Val Lys Gly Ala Pro Glu Gly Val Leu Asp Arg Cys Thr His Val Arg Val Gly Thr Lys Lys Val Pro Met Thr   189
488 ATG TTT GTA AAA GGA GCC CCT GAA GGT GTC CTT GAC CGT TGC ACA CAT GTC CGT GTA GGA ACT AAA AAG GTC CCA ATG ACT
   Pro Ala Ile Met Asp Lys Ile Leu Glu Val Thr Arg Ala Tyr Gly Thr Gly Arg Asp Thr Leu Arg Cys Leu Ala Leu Ala   216
569 CCA GCT ATT ATG GAC AAG ATT TTG GAA GTC ACT CGT GCT TAT GGT ACT GGG CGA GAT ACC CTT CGA TGC CTC GCT TTG GCC
   Thr Ile Asp Asp Pro Met Asp Pro Lys Asp Met Asp Ile Ile Asp Ser Thr Lys Phe Val Lys Tyr Glu Gln Asn Cys Thr   243
650 ACG ATT GAC GAT CCA ATG GAC CCT AAA GAC ATG GAC ATT ATT GAT TCT ACC AAA TTT GTC AAA TAT GAG CAA AAC TGT ACA
   Phe Val Gly Val Val Gly Met Leu Asp Pro Pro Arg Lys Glu Val Leu Asp Ala Ile Glu Arg Cys Arg Ala Ala Gly Ile   270
731 TTT GTT GGG GTT GTT GGT ATG CTT GAT CCC CCT CGT AAA GAA GTC CTT GAT GCA ATT GAA CGC TGT CGT GCT GCT GGT ATT
   Arg Val Ile Val Ile Thr Gly Asp Asn Lys Ala Thr Ala Glu Ala Ile Cys Arg Arg Ile Gly Val Phe Gly Glu Asp Glu   297
812 CGT GTC ATT GTT ATT ACT GGC GAC AAC AAG GCC ACC GCT GAA GCT ATC TGT AGG CGA ATT GGT GTC TTT GGT GAA GAT GAA
   Asn Thr Glu Gly Met Ala Tyr Thr Gly Arg Glu Phe Asp Asp Leu Ser Val Glu Gly Gln Arg Asp Ala Val Ala Arg Ser   324
893 AAT ACG GAA GGT ATG GCC TAC ACT GGA CGC GAA TTC GAT GAC TTG TCT GTT GAA GGT CAA CGT GAT GCA GTT GCT CGT TCT
   Arg Leu Phe Ala Arg Val Glu Pro Phe His Lys Ser Lys Ile Val Glu Tyr Leu Gln Gly Met Gly Glu Ile Ser Ala Met   351
974 CGT CTT TTC GCT CGC GTT GAG CCT TTC CAC AAA TCC AAG ATT GTT GAG TAT TTA CAA GGC ATG GGT GAA ATT TCC GCT ATG
   Thr Gly Asp Gly Val Asn Asp Ala Pro Ala Leu Lys Lys Ala Glu Ile Gly Ile Ala Met Gly Ser Gly Thr Ala Val Ala   378
1055 ACT GGT GAC GGT GTG AAC GAT GCG CCA GCT TTG AAA AAA GCC GAA ATT GGT ATC GCC ATG GGG TCT GGC ACT GCT GTC GCC
   Lys Ser Ala Ala Glu Met Val Leu Ala Asp Asp Asn Phe Ser Thr Ile Val Ala Ala Val Glu Glu Gly Arg Ala Ile Tyr   405
1136 AAA TCT GCC GCT GAG ATG GTT TTA GCT GAT GAT AAC TTC TCT ACC ATC GTT GCC GCC GTT GAA GAA GGT CGA GCT ATC TAC
   Asn Asn Met Lys Gln Phe Ile Arg Tyr Leu Ile Ser Ser Asn Ile Gly Glu Val Val Ser Ile Phe Leu Thr Ala Ala Leu   432
1217 AAC AAC ATG AAA CAG TTT ATC CGC TAC TTG ATC TCT TCC AAT ATT GGT GAA GTT GTC TCT ATT TTC TTG ACT GCT GCC CTC
   Gly Leu Pro Glu Ala Leu Ile Gln Val Gln Leu Leu Trp Val Asn Leu Val Thr Asp Gly Leu Pro Ala Thr Ala Leu Gly   459
1298 GGA CTC CCT GAA GCT TTG ATT CCA GTA CAA CTT TTG TGG GTC AAT TTG GTC ACT GAT GGT CTC CCA GCT ACT GCT CTT GGT
   Phe Asn Pro Pro Asp Leu Asp Ile Met Asn Lys Pro Pro Arg Arg Ala Asp Glu Gly Leu Ile Thr Gly Trp Leu Phe Phe   486
1379 TTC AAT CCA CCC GAT CTT GAC ATT ATG AAC AAA CCA CCC CGT CGT GCT GAT GAG GGT CTC ATT ACA GGA TGG CTG TTT TTC
   Arg Tyr Met Ala Ile Gly Thr Tyr Val Gly Ala Ala Thr Val Gly Ala Ala Ala His Trp Phe Met Met Ser Pro Thr Gly   513
1460 CGT TAT ATG GCT ATT GGT ACA TAT GTT GGT GCT GCC ACT GTA GGT GCT GCC GCG CAT TGG TTC ATG ATG TCA CCC ACT GGA
   Pro Gly Leu Asn Phe Tyr Gln Leu Ser His His Leu Gln Cys Thr Pro Glu Asn Glu Tyr Phe Glu Gly Ile Asp Cys Glu   540
1541 CCC GGT CTT AAT TTC TAT CAG TTG AGC CAT CAC CTT CAA TGT ACA CCT GAG AAT GAA TAT TTT GAA GGT ATT GAT TGC GAA
   Ile Phe Ser Asp Pro His Pro Met Met Ala Leu Ser Val Leu Val Thr Ile Glu Met Leu Asn Ala Ile Asn Ser Leu   567
1622 ATA TTT AGT GAC CCT CAT CCA ATG ACT ATG GCT TTG TCA GTA TTG GTC ACT ATT GAG ATG TTG AAT GCC ATT AAC AGC TTG
   Ser Glu Asn Gln Ser Leu Leu Val Met Pro Pro Trp Ser Asn Ile Trp Leu Ile Ser Ala Ile Cys Leu Ser Met Thr Leu   594
1703 TCT GAG AAC CAG TCT CTC TTG GTA ATG CCA CCG TGG TCA AAT ATC TGG CTC ATC TCT GCC ATC TGC CTT TCC ATG ACT CTT
   His Phe Val Ile Leu Tyr Val Glu Ile Leu Ser Thr Val Phe Gln Ile Cys Pro Leu Thr Leu Thr Glu Trp Ile Val Val   621
1784 CAC TTT GTT ATC CTT TAT GTC GAA ATA CTG TCT ACT GTC TTC CAA ATT TGT CCG CTG ACA TTA ACA GAA TGG ATT GTT GTA
   Leu Lys Ile Ser Phe Pro Val Leu Leu Leu Asp Glu Val Leu Lys Phe Val Ala Arg Lys Tyr Thr Asp Gly Met Pro Leu   648
1865 TTG AAG ATC AGT TTC CCA GTA TTG CTT CTA GAT GAA GTT CTC AAA TTC GTA GCC AGA AAA TAT ACC GAT G/GT ATG CCG CTT
   Ser Ser Tyr Phe Val Asp Ala Trp Gly Leu Val Leu Ala Trp Ala Leu Phe Phe Gly Val Ile Phe Tyr Ser Pro Leu Stop   674
1946 AGT TCA TAT TTT GTT GAT GCA TGG GGT CTC GTC CTC GCA TGG GCC TTA TTT TTT GGA GTG ATA TTC TAT TCC CCG TTA TAA
2027 CTATGAACGATTTTAGGAACTCTATGTACCCCTTTGAGACGTGCGACGAAAGTAGTATTGCTCTCTGAGAACATGGCAACAAATTCTTTTGAAGTTTTTAATTATTT
2134 ATTTAAAATGTTTTTTCTCTAGTGACGGGAACAAATGAACACATTAATGCCATCTATCACAAAATTTGTGCAATTGTCAAAATTTCTAAGATATAAATGTTATTGAA
2241 ATACATGTTAGAAATTTACATTCCGTTGGTGGTGAATAATAAACTTCGCCCCAACCATTTATATGTATACCTTTCCACAAATTCTCATTTTCTCAAAAACTATGCGG
2348 AAAATTATATTGAAGTAATTAAATTGATAAATTATATTGAAGTAATTAAATTGATAAATTTCTTTATTAATTGAAAAGATGTAGAAATACTGAAATTTTCCTTTCC
2455 TTATTTGTAAAAATGCCAAATAGCTATATGTTTATCATTTTCTTCTTTAAAGTCATCGGTAATTTGCAACACAAGAAAATTTGTGGAACGGTCTGTATTTCAGTTTC
2562 AAAATGAACTCTTTTCTTACGACTGTGTTGAAGTGTATTTATAAAACCTAGCATGTATTTGTCTTGTCTGTTTTATATAGAAAACCCCATGTTTTGAAATAAAAGTT
2669 AGTTTTTGCGTT(A)n
```

FIG. 1. **Restriction map of the cDNA clones coding for *A. franciscana* SERCA and nucleotide sequence of the clone pArATCa604.** *Panel A*, the restriction map of the cDNA clones isolated from a cDNA library from cryptobiotic embryos (clones pArATCa501, -505, -508, and -517) is shown together with the map of the cDNA clones previously isolated (pArATCa7, -151, and -412) and that of a clone isolated from a 20-h-old nauplii cDNA library using the clone pArATCa501 as probe (clone pArATCa604). Restriction maps were determined by single and double digestions with the indicated enzymes. The clones have been grouped in two classes according to their restriction maps. The *box* indicates the protein coding regions of each class of cDNA clones and the *vertical dashed line* the point of divergence of the two

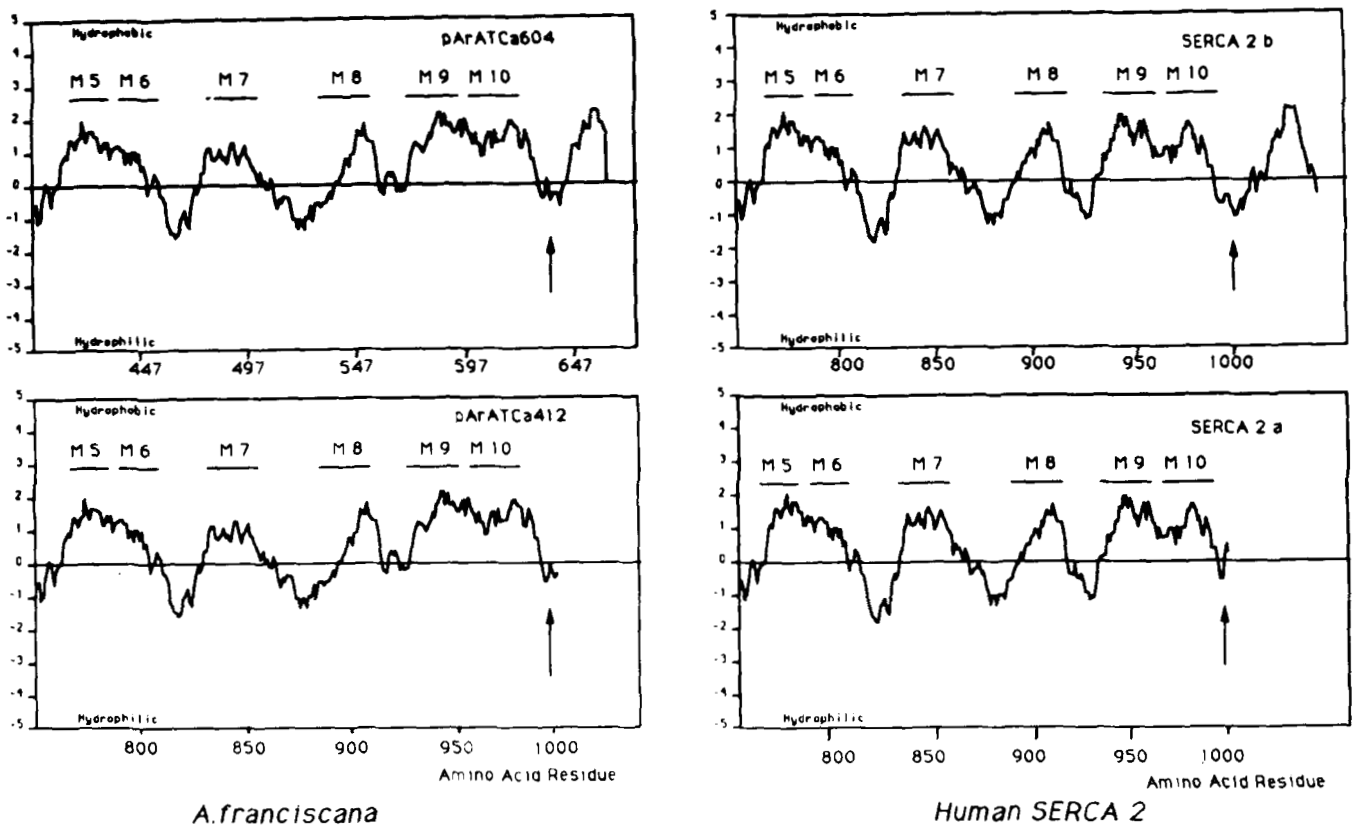


FIG. 2. **Hydropathicity plot of *A. franciscana* SERCA isoforms in comparison with that of the human SERCA2 isoforms.** The hydropathicity plot of the C-terminal region of *A. franciscana* SERCA isoforms (pArATCa604 and -412) has been determined by the method of Kyte and Doolittle using the Macintosh program GeneJockey (Biosoft). Putative transmembrane domains, numbered according to the model of Brandl *et al.* (1986), are indicated (*M5–M10*). The hydropathicity plots of the human SERCA2a and SERCA2b isoforms, determined with the same program, are shown for comparison. Human sequences were taken from Lytton and MacLennan (1988). The *arrow* indicates the divergence point of the two isoforms in each species.

coding for the 3′ end of the gene. The location of these exons on the genomic clones and the nucleotide sequence of their borders are shown in Fig. 4.

The most 5′ exon (*exon A* in Fig. 4) codes for a region common to both types of cDNA clones and that extends from the amino acid 919 to the amino acid 958 of the protein (nucleotides 1699–1816 of the pArATCa604 clone). The limits of this exon match perfectly with those of the exon 21 of the rabbit gene SERCA1 (Korczak *et al.*, 1988) and of the homologous exon of the gene SERCA2 (Lytton and MacLennan, 1988). The intermediate exon (*exon B′*) codes for the common amino acids 959–997 (nucleotides 1818–1934 of pArATCa604) and from there until the end is identical to clone pArATCa604. The most 3′-terminal exon (exon C) codes for the sequence that is specific for the 3′ end of the clone pArATCa412 (amino acids 998–1003 and 3′-untranslated sequence).

The putative donor and acceptor splicing sites of the three exons are shown in Fig. 4. All these sequences are highly similar to the consensus sequences defined for other species (Mount, 1982). The 3′ end exon specific for the cDNA clone pArATCa412 (exon C) is preceded by a consensus acceptor sequence. The corresponding donor sequence can be recognized at the end of the region common to both cDNAs in the penultimate exon (exon B). These data show that the 4.5-kb mRNA is originated by a splicing that uses a donor sequence internal to the 5.2-kb mRNA. The 5.2-kb mRNA do not have this last exon, since it uses a polyadenylation site located before the last exon of the gene. The alternative splicing found in *Artemia*, described for the first time in invertebrates, is similar to the one described in the vertebrate SERCA2 gene (Lytton and MacLennan, 1988).

DISCUSSION

The data presented in this paper show the existence of great similarity between the characterized *Artemia* SERCA gene and the gene SERCA2 from mammals and birds. All these genes code for two isoforms of the enzyme originated by similar alternative splicing events. Furthermore, the two isoforms generated from these genes differ solely in their C-terminal amino acids, and one of them presents a extension of hydrophobic amino acids with the characteristics of a potential transmembrane domain. The conservation of this extra transmembrane domain between vertebrates and crustaceans suggests that it may play an important role in the function of this enzyme in the cells where this isoform is

classes of cDNA clones as determined by nucleotide sequencing (see below). *Panel B*, nucleotide sequence of the clone pArATCa604 and the predicted amino acid sequence of the encoded protein. The nucleotide sequence of the clone pArATCa604 was determined from both DNA strands. It is almost identical to that of other cDNA clones previously sequenced up to the nucleotide 1934 (indicated by a *box* in *panel B* and by a *dashed line* in *panel A*) and is different from previous clones onwards. Nucleotides different from the other cDNA clones in the common region are *underlined*. The only amino acid that is different from the other clones in the common region is also *underlined*.

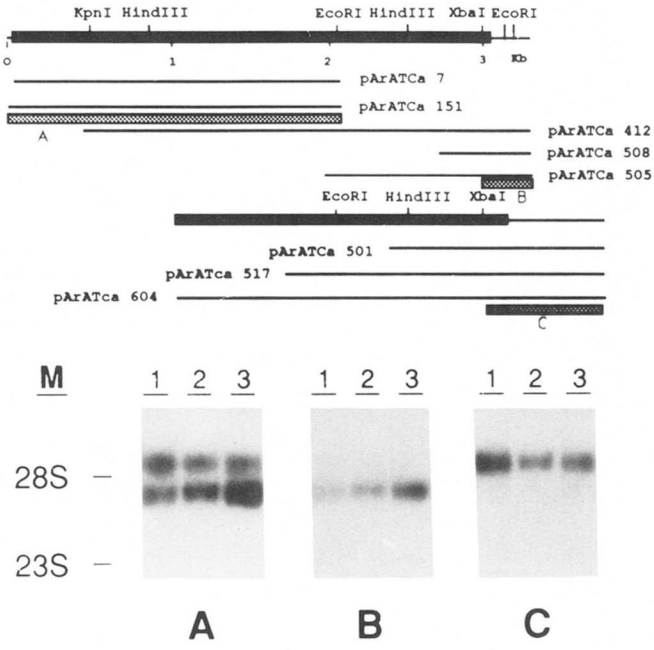


FIG. 3. **Northern blot analyses of the mRNAs encoded by *A. franciscana* SERCA isoforms.** Fifteen micrograms of total RNA from *A. franciscana* embryos cultured for 6 (*lanes 1*), 8 (*lanes 2*), or 10 h (*lanes 3*) were electrophoresed in 2.2 M formaldehyde, 1.5% agarose gels, transferred to nylon membranes, and hybridized with the following probes: *A*, the *Eco*RI fragment of the clone pArATCa151, indicated by *box A* underneath the restriction map of this clone; *B*, the 283-base pair *Xba*I-*Xho*I fragment of the clone pArATCa505, specific for this class of cDNA clones and indicated by *box B* underneath their restriction map; *C*, the 777-base pair *Xba*I-*Xho*I fragment of the clone pArATCa501, specific for this class of cDNA clones and indicated by *box C*.

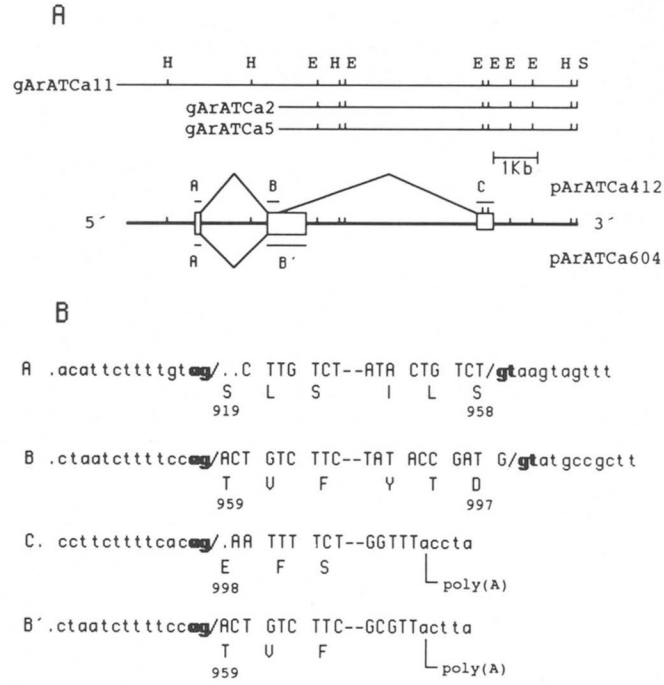


FIG. 4. **Structure of the C-terminal region of *A. franciscana* SERCA gene.** Genomic clones were isolated that hybridized to the C-terminal region of the *A. franciscana* SERCA gene using the cDNA probes B and C described in Fig. 3. *Panel A* shows the restriction map of the three genomic clones obtained (gArATCa2, -5, and -11) as determined by digestion with *Eco*RI (*E*), *Sal*I (*S*), *Hin*dIII (*H*), or combinations of these enzymes. The location of cDNA coding regions was determined by hybridization and nucleotide sequencing and is indicated by *open boxes*. In *panel B* are indicated the nucleotide sequences and location of the boundaries of each identified exon. The amino acids encoded by each exon are numbered according to the complete amino acid sequence of the protein (Palmero and Sastre, 1989). Exon A is common to the two cDNA isoforms and is bordered by consensus acceptor and donor splicing sequences. The absolutely conserved acceptor AG and donor GT nucleotides are indicated in *boldface type* in all the intron/exon boundaries. The 5′ region of exon B is also common to both isoforms. In one of the isoforms, a donor splicing site is recognized after this common region, which is fused to exon C. The acceptor and donor sequences used in this splicing are shown in exons B and C. Exon C ends at the polyadenylation site indicated as *poly(A)*. Alternatively, exon B can be ended at a polyadenylation site located before exon C to give exon B′ that codes for the other isoform of the protein. The acceptor splicing site and polyadenylation site of exon B′ are also indicated.

expressed. We are studying at the present time if the expression of the two isoforms generated in *Artemia* are also tissue-specific, as described in mammals.

Little is known about the biological significance of the existence of the two isoforms. Campbell *et al.* (1992) recently reported that the C-terminal extension of the SERCA2b isoform of birds expands the endoplasmic reticulum membrane so that, whereas the SERCA2a isoform has 10 transmembrane domains, the SERCA2b isoform has 11 transmembrane domains. This structural difference does not seem to change the enzymatic properties of the protein (Campbell *et al.*, 1991; Lytton *et al.*, 1991). It is possible that this extra transmembrane domain could change the regulatory properties of the enzyme or its interactions with other cellular components. It is interesting to notice a difference between this extra transmembrane domain and the other common transmembrane domains; the amino acid sequence of the common transmembrane domains is highly conserved between *Artemia* and mammals (Palmero and Sastre, 1989), whereas the amino acid sequence of the extra transmembrane domain is poorly conserved. The conservation of the common transmembrane domains is probably due to the important function of these domains in calcium binding and transport, as shown by *in vitro* mutagenesis (Clarke *et al.*, 1990). The poor conservation of the extra transmembrane domain could be in agreement with the idea that it is not involved in the enzymatic activity of the protein but in the regulatory interactions with other cellular components that might have diverged between *Artemia* and mammals.

The existence of similar splicing events to originate homologous isoforms between crustaceans and vertebrates also has some interesting evolutionary implications. The simplest explanation for the conservation of the alternative splicing would be that the precursor gene for vertebrate SERCA2 genes and this *Artemia* gene already coded for two isoforms generated by alternative splicing of the primary transcript. An alternative hypothesis, which we think is less probable, would be that mammalian and bird SERCA2 genes and the *Artemia* gene generated independently similar alternative splicing mechanisms during evolution.

Other evidence suggests that the three vertebrate SERCA genes, the *Drosophila* genes, and the *Artemia* genes have a common ancestor. Vertebrate genes are more homologous to one another (75–84%) than to the invertebrate genes (67–71% to the *Drosophila* gene, Magyar and Váradi (1990); and 67–71% to the Artemia gene, Palmero and Sastre (1989)). The invertebrate genes are unique genes (Magyar and Váradi, 1990)[2] and present a similar homology with the three vertebrate genes (Magyar and Váradi, 1990; Palmero and Sastre, 1989). Altogether, these data suggest the existence of a unique

[2] R. Escalante and L. Sastre, unpublished results.

SERCA ancestor gene that gave rise to the three genes present in birds and mammals and to the single invertebrate genes. This ancestor gene should have coded for the two isoforms of the protein originated by the alternative splicing that has been conserved in vertebrate SERCA2 genes and in the *Artemia* gene. This alternative splicing would have been lost in vertebrate genes SERCA1 and SERCA3 during evolution. This hypothesis would predict that the same alternative splicing will be found in other invertebrate genes for which no data is available. The study of the SERCA genes in other species would be necessary to test these hypothesis.

Acknowledgments—We thank Drs. Rafael Garesse, Rosario Perona, and Jaime Renart for critical reading of the manuscript and Antonio Fernandez and Elena Fraga for the preparation of the figures.

REFERENCES

Brandl, C. J., Green, N. M., Korczak, B., and MacLennan, D. H. (1986) *Cell* **44,** 597–607
Brandl, C. J., de Leon, S., Martín, D. R., and MacLennan, D. H. (1987) *J. Biol. Chem.* **262,** 3768–3774
Burk, S. E., Lytton, J., MacLennan, D. H., and Shull, G. E. (1989) *J. Biol. Chem.* **264,** 18561–18568
Campbell, A. M., Kessler, P. D., Sagara, Y., Inesi, G., and Fambrough, D. M. (1991) *J. Biol. Chem.* **266,** 16050–16055
Campbell, A. M., Kessler, P. D., and Fambrough, D. M. (1992) *J. Biol. Chem.* **267,** 9321–9325
Chen, E. J., and Seeburg, P. H. (1985) *DNA* **4,** 165–170
Clarke, D. M., Loo, T. W., and MacLennan, D. H. (1990) *J. Biol. Chem.* **265,** 6262–6267
Cruces, J., Sebastián, J., and Renart, J. (1981) *Biochem. Biophys. Res. Commun.* **98,** 404–409
Gunteski-Hamblin, A.-M., Greeb, J., and Shull, G. E. (1988) *J. Biol. Chem.* **263,** 15032–15040
Korczak, B., Zarain-Herzberg, A., Brandl, C. J., Ingles, C. J., Green, N. M., and MacLennan, D. H. (1988) *J. Biol. Chem.* **263,** 4813–4819
Lytton, J., and MacLennan, D. H. (1988) *J. Biol. Chem.* **263,** 15024–15031
Lytton, J., Zarain-Herzberg, A., Periasamy, M., and MacLennan, D. H. (1989) *J. Biol. Chem.* **264,** 7059–7065
Lytton, J., Westlin, M., and Hanley, M. R. (1991) *J. Biol. Chem.* **266,** 17067–17071
Magyar, A., and Váradi, A. (1990) *Biochem. Biophys. Res. Commun.* **173,** 872–877
Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472
Osuna, C., and Sebastián, J. (1980) *Eur. J. Biochem.* **109,** 383–389
Palmero, I., and Sastre, L. (1989) *J. Mol. Biol.* **210,** 737–748
Sherman, F., Fink, G. R., and Hicks, J. B. (1986) *Laboratory Course Manual for Methods in Yeast Genetics,* pp. 143–144, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY